Coffee quality

T.sriram prasath

60524OLE003

B3MAY(online)

Introduction

The coffee quality dataset in Python is often used for analyzing and predicting coffee characteristics. This dataset typically includes attributes like aroma, flavor, acidity, body, and balance, along with overall quality ratings. It can be utilized to build machine learning models that predict quality scores based on these features. Data scientists use this dataset to perform exploratory data analysis (EDA), feature engineering, and model evaluation to understand the factors influencing coffee quality. Popular Python libraries for handling such datasets include Pandas for data manipulation and Scikit-learn for building predictive models.

Aim

To check coffee quality using Python, you can analyze various attributes such as aroma, flavor, acidity, and body. Start by collecting data on these attributes through sensory evaluation forms. Use Python libraries like Pandas for data management and Scikit-learn for building predictive models. For instance, you can implement machine learning algorithms to classify coffee quality based on sensory attributes. Visualize the results with Matplotlib or Seaborn to identify trends and patterns. Additionally, integrate natural language processing (NLP) if dealing with descriptive feedback. This approach helps in systematically assessing and improving coffee quality.

Problem statement

We want to evaluate the quality of coffee based on three factors: acidity, bitterness, and aroma. Each factor is rated on a scale from 0 to 10, with 10 being the best score. The quality of the coffee is determined using the following criteria:

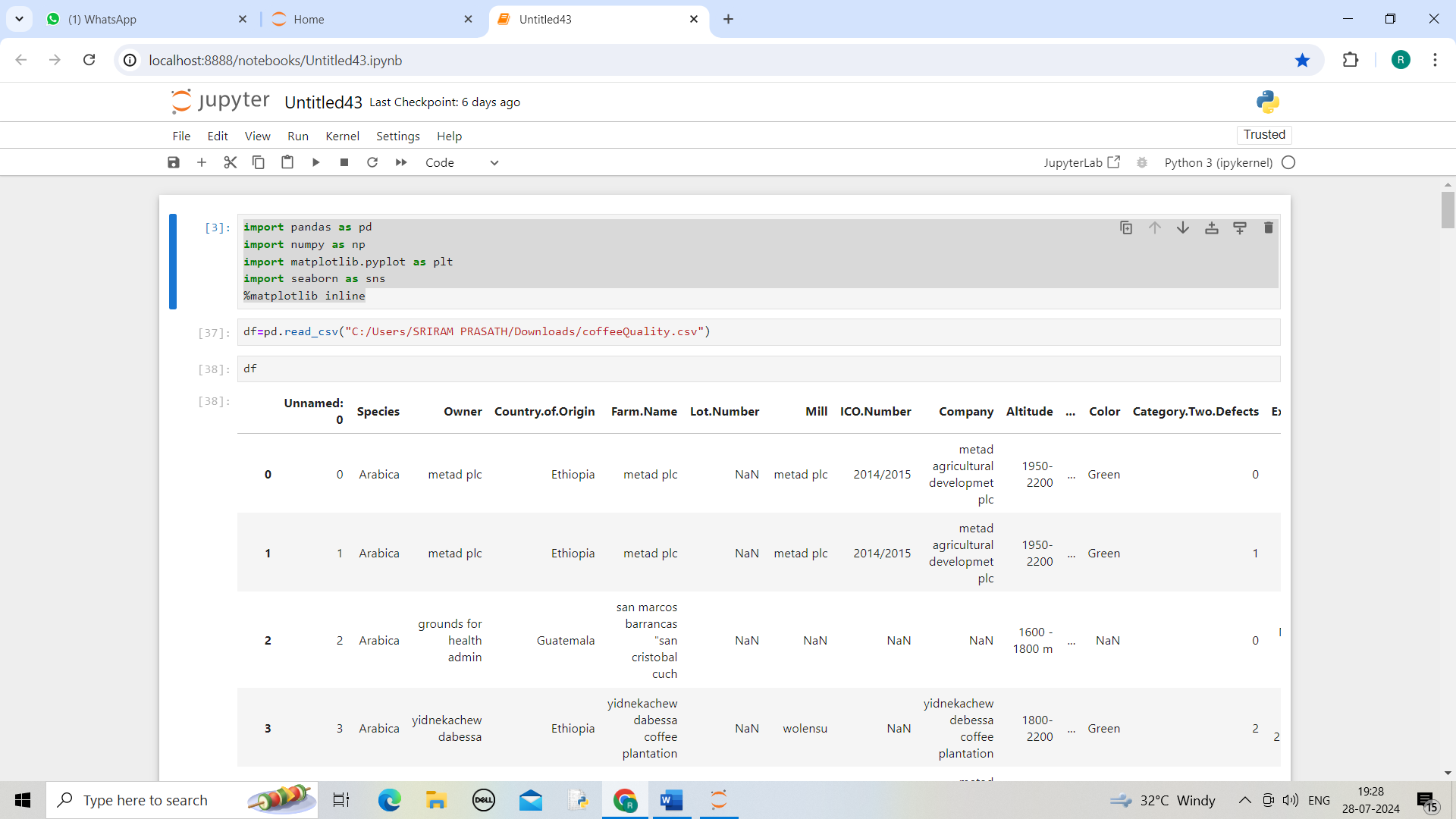
Acidity: A score of 7 or higher is considered good.

Bitterness: A score of 5 or lower is considered acceptable.

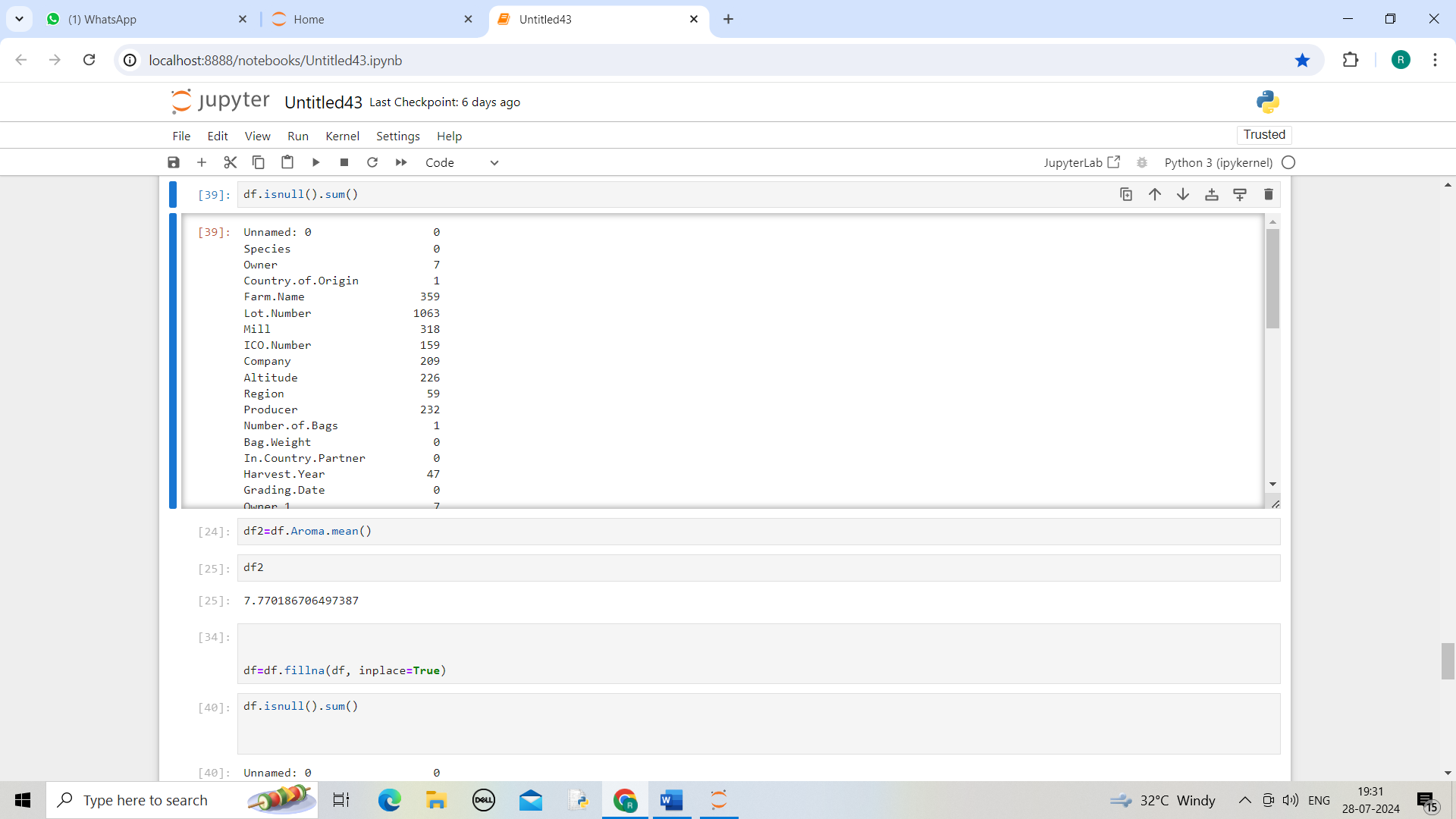
Aroma: A score of 8 or higher is considered excellent.

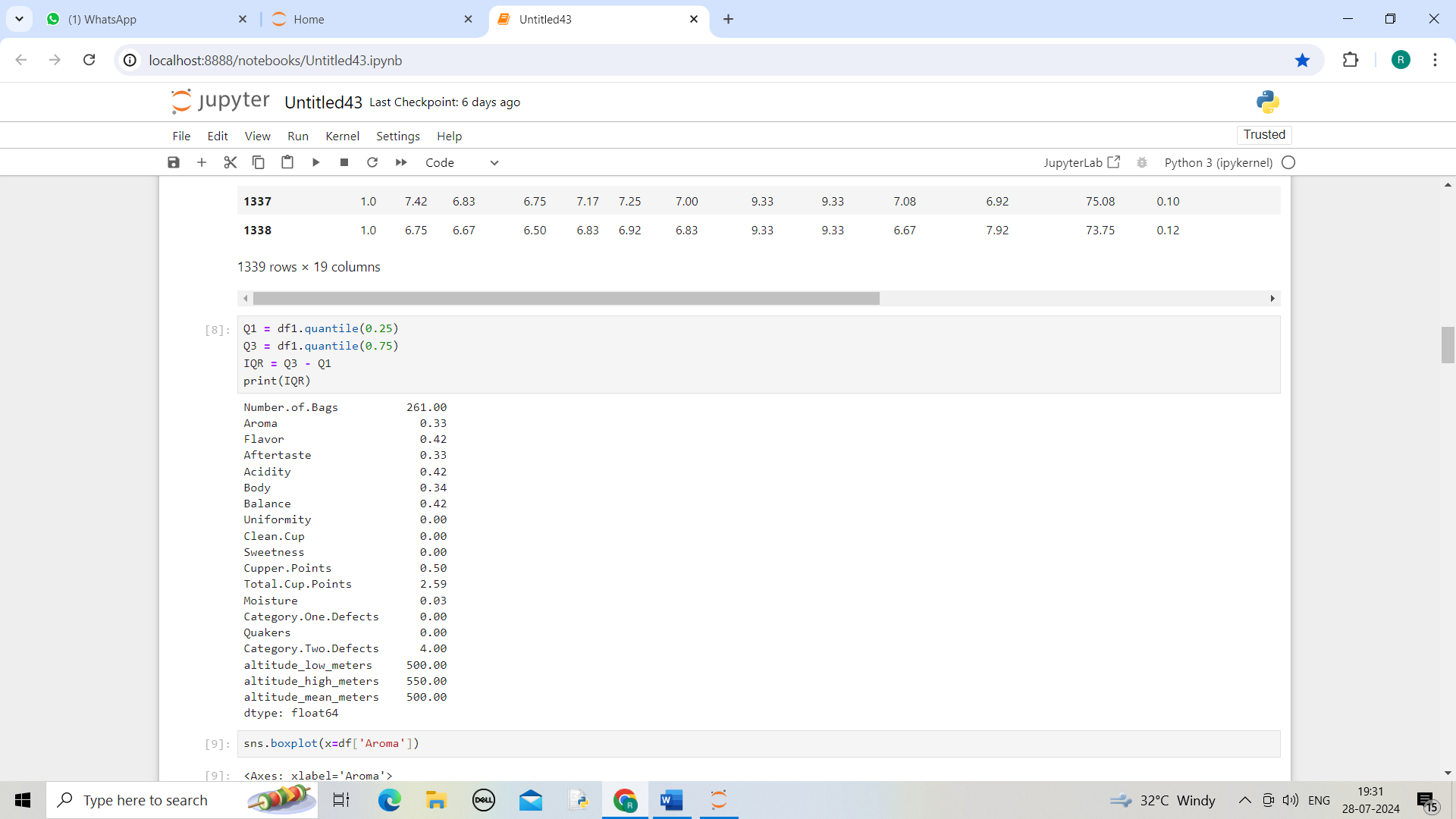
To assess the overall quality of the coffee, we'll use these criteria to generate a quality rating.

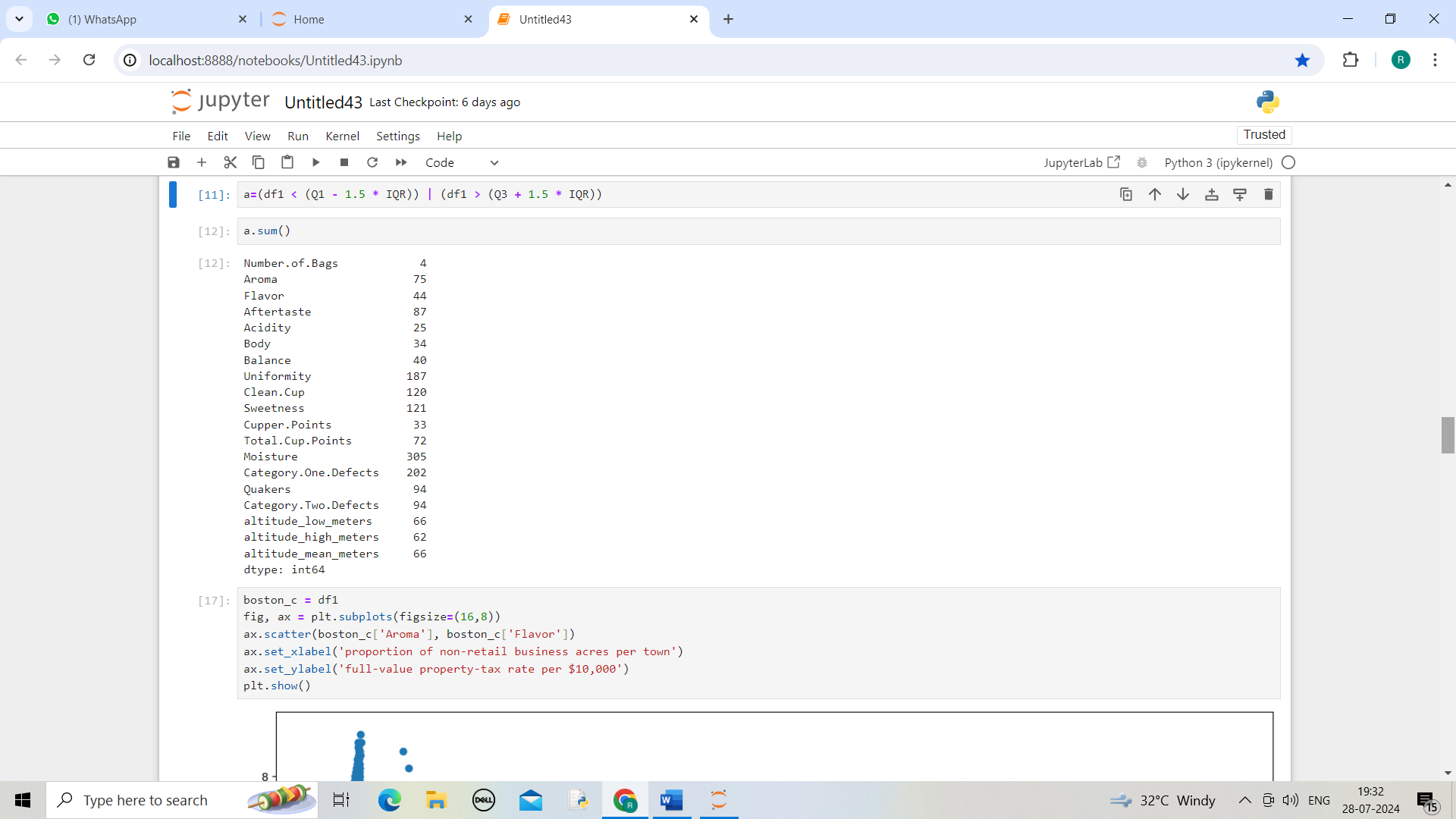
Data understanding



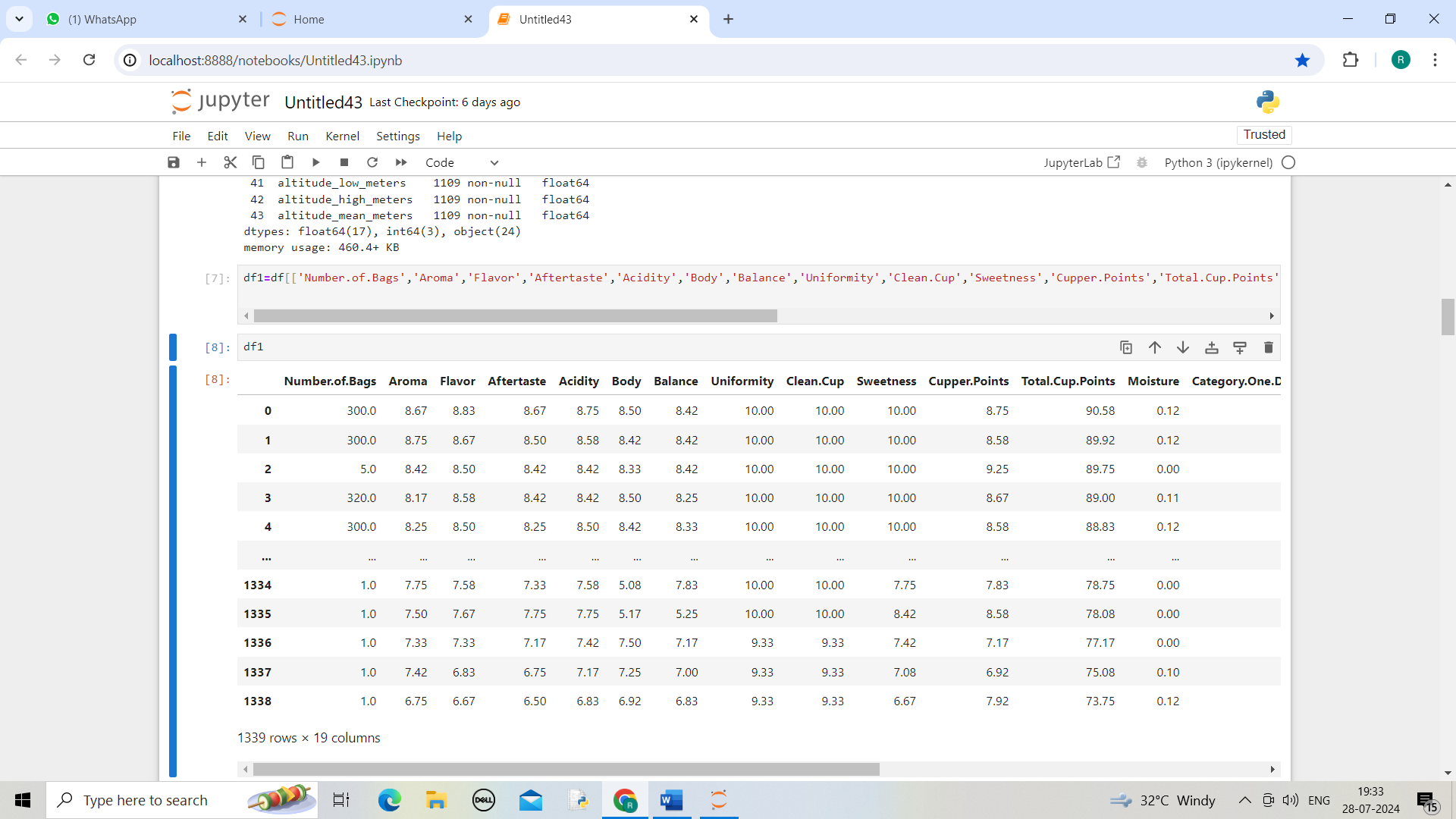
Data cleaning



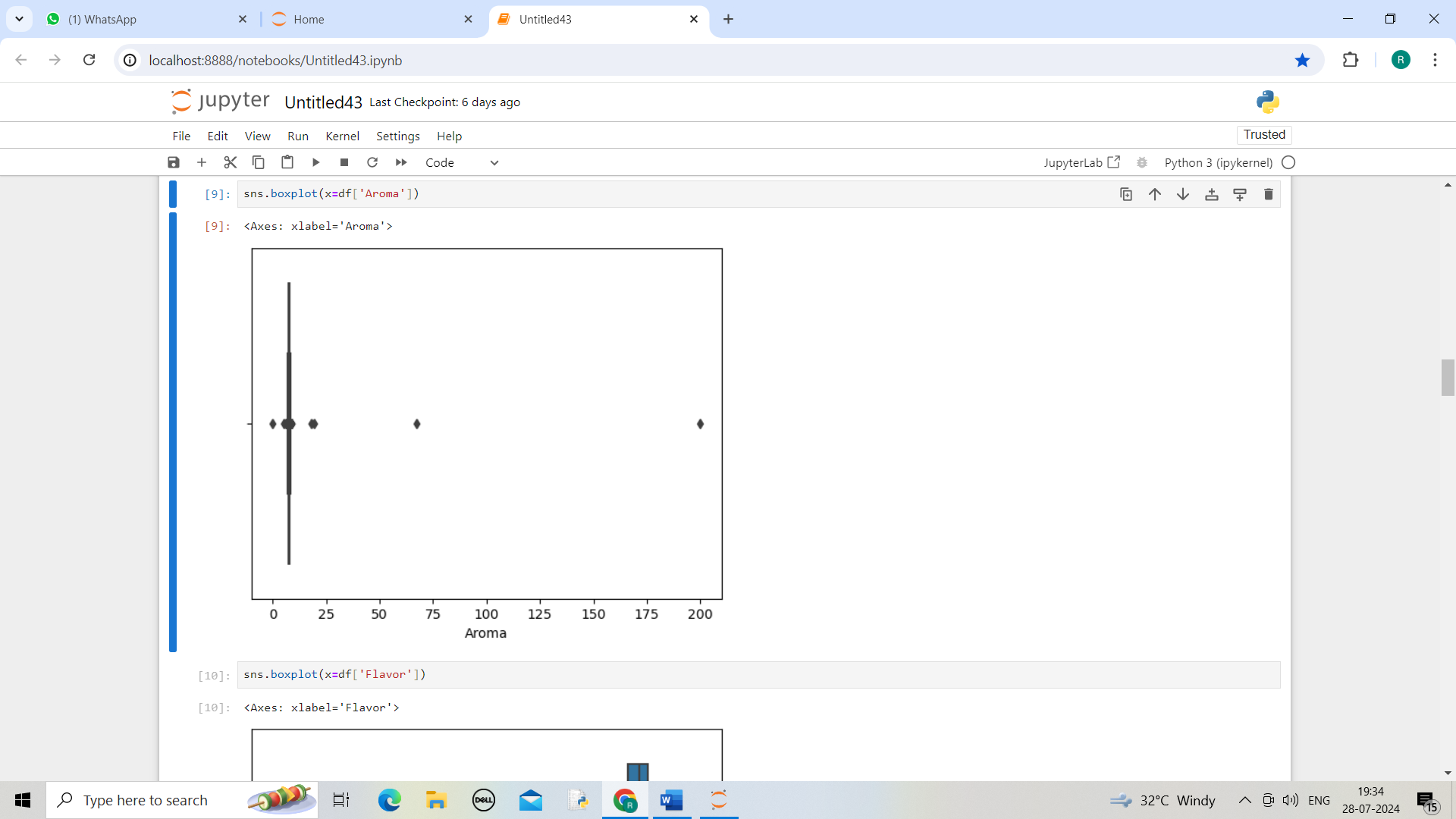


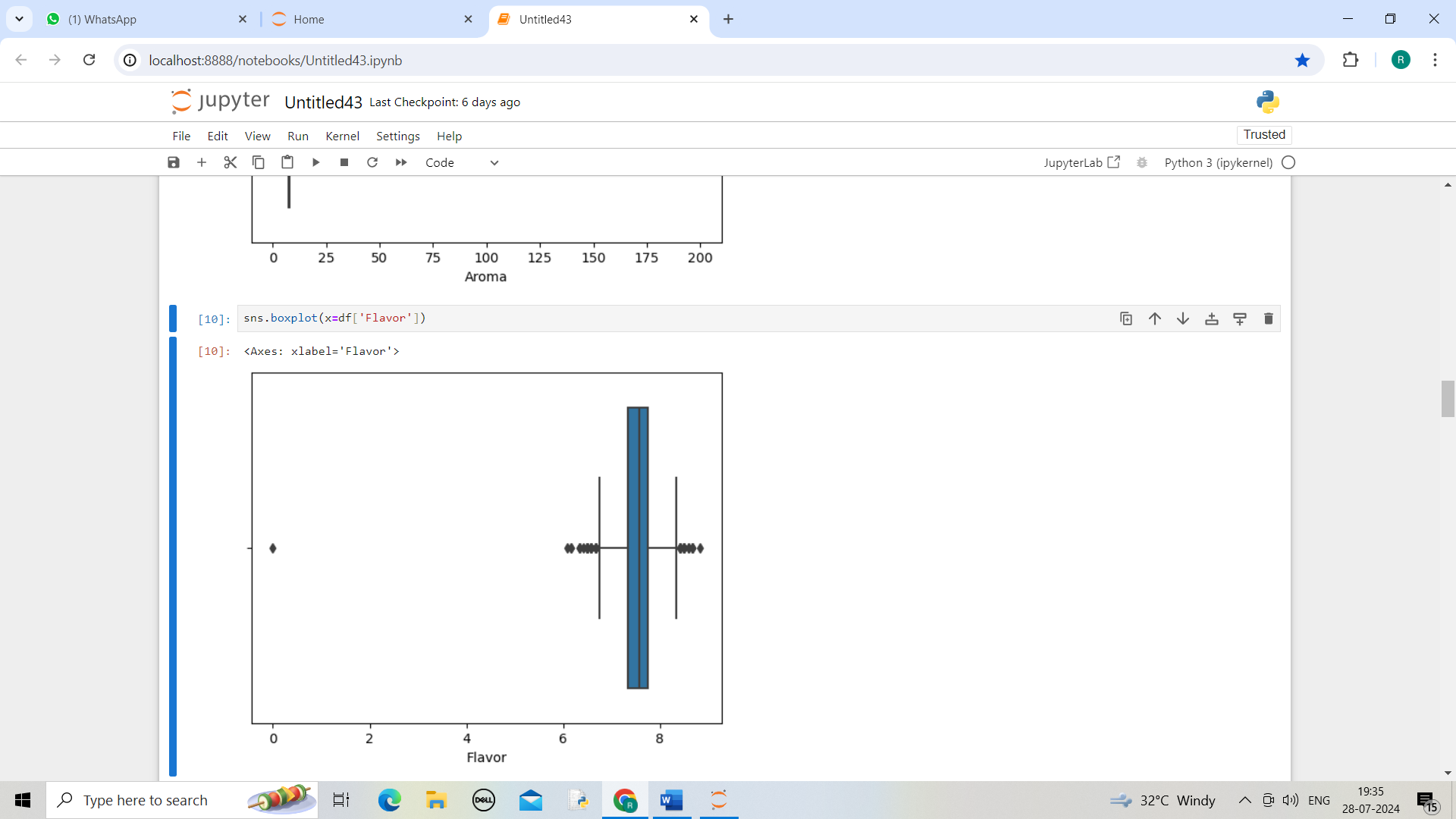


Filtering Data



EDA-Univariet analysis





Bivariate Analysis

Bivariate analysis in coffee quality checking involves examining the relationship between two variables to understand how they influence each other. For instance, you might analyze the correlation between roast level and flavor profile or between bean size and cup acidity. This type of analysis helps identify patterns or trends that can inform decisions on improving coffee quality. Key steps in bivariate analysis include:

Data Collection: Gather data on the variables of interest (e.g., roast time, flavor notes, bean origin).

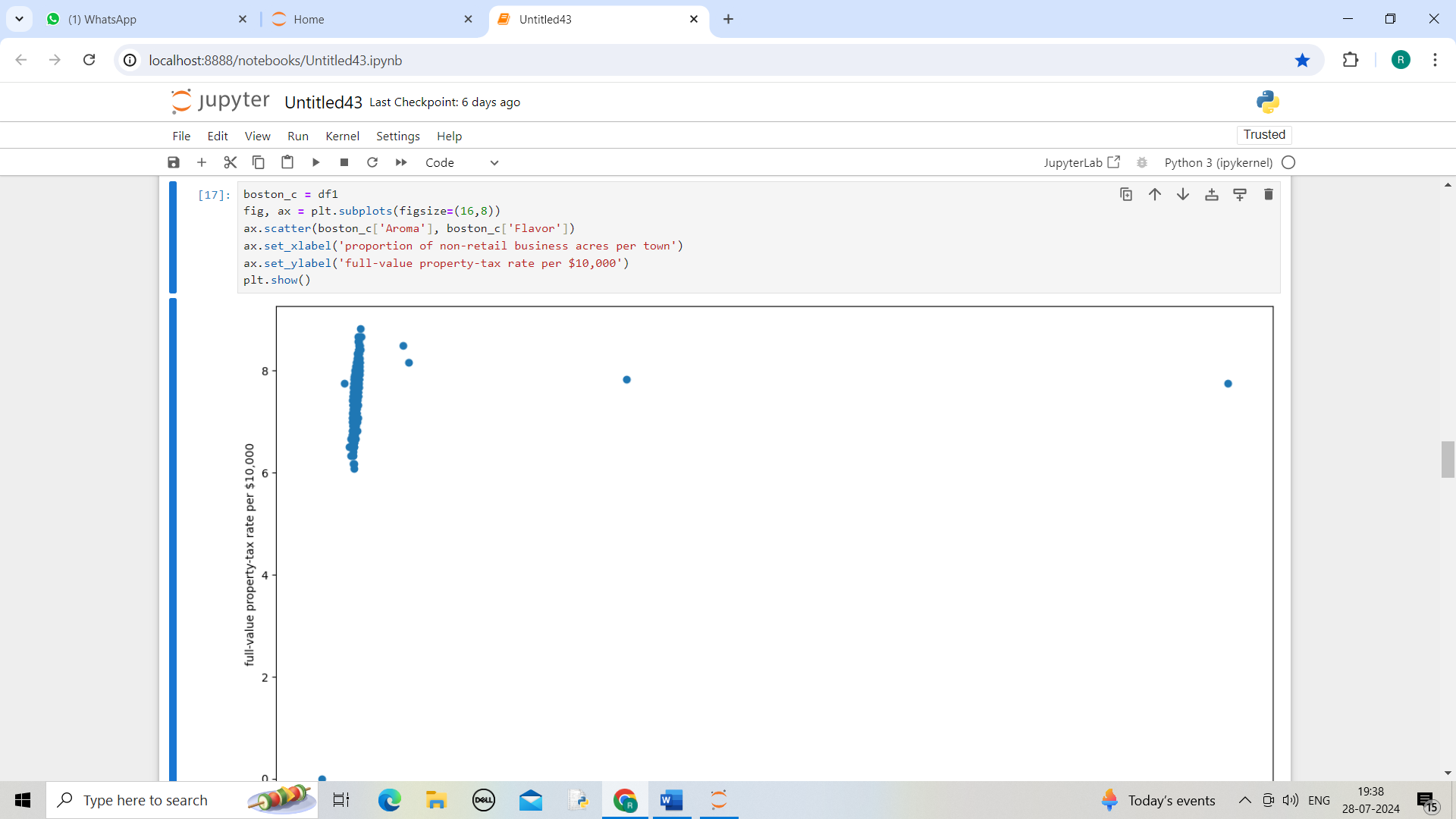
Data Preparation: Clean and preprocess the data to ensure accuracy.

Exploratory Data Analysis (EDA): Use scatter plots, correlation coefficients, or other techniques to visualize and quantify relationships.

Statistical Testing: Apply statistical methods (e.g., regression analysis) to test the strength and significance of the relationships.

Interpretation: Draw conclusions from the analysis to make informed decisions about coffee quality.

This approach helps in understanding how different factors contribute to the overall quality and can guide adjustments to improve the final product.



Multivariet Analysis

Bivariate analysis in coffee quality checking without coding can be conducted through manual methods and statistical tools like spreadsheets or specialized software. Here's a guide on how to perform this analysis:

1. Identify Variables

Determine the two variables you want to analyze. For example:

- Roast level vs. flavor profile

- Bean size vs. acidity

2. Collect Data

Gather data on these variables. This could be from sensory evaluations, lab tests, or other quality assessments.

3. Create a Data Table

Enter the data into a table or spreadsheet. For instance:

| Roast Level | Flavor Profile |

|-------------|----------------|

| 1 | 80 |

| 2 | 85 |

| 3 | 87 |

| 4 | 83 |

| 5 | 90 |

4. Visualize the Data

Use charts to visualize the relationship between the two variables:

Scatter Plot: Plot one variable on the x-axis and the other on the y-axis to observe if there's a pattern or correlation.

How to create a scatter plot in Excel:

-Select your data.

- Go to the "Insert" tab.

- Choose "Scatter" from the Charts section.

5. Calculate Correlation

Determine if there’s a relationship between the variables by calculating the correlation coefficient.

Using Excel:

Use the CORREL function: =CORREL(array1, array2), where array1 and array2 are the ranges of your two variables.

Interpreting Correlation:

Positive Correlation: As one variable increases, the other tends to increase.

Negative Correlation: As one variable increases, the other tends to decrease.

No Correlation: No discernible pattern between the variables.

6. Interpret Results

Review the scatter plot and correlation coefficient:

Strong Positive Correlation: Points cluster around an upward slope.

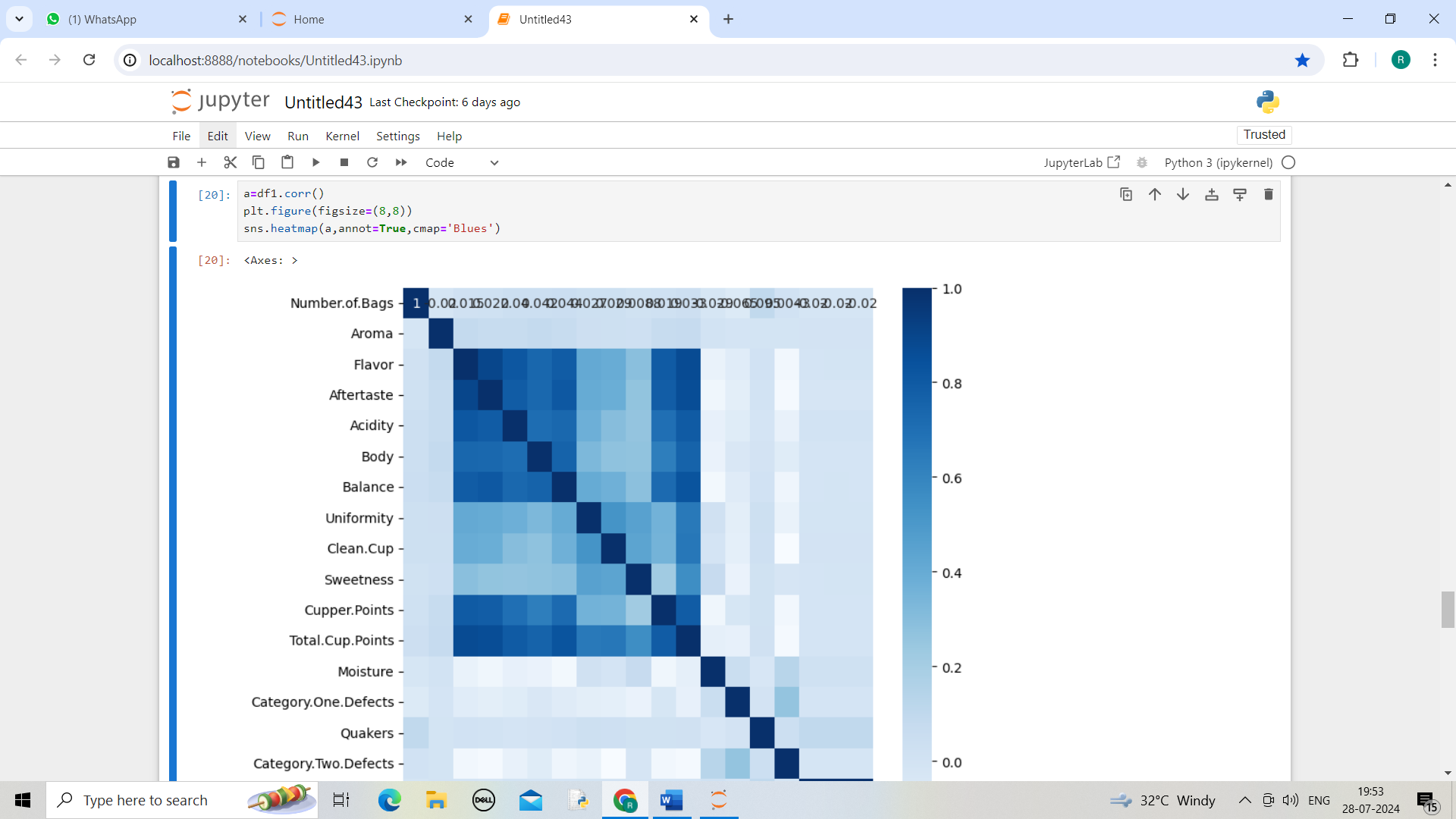
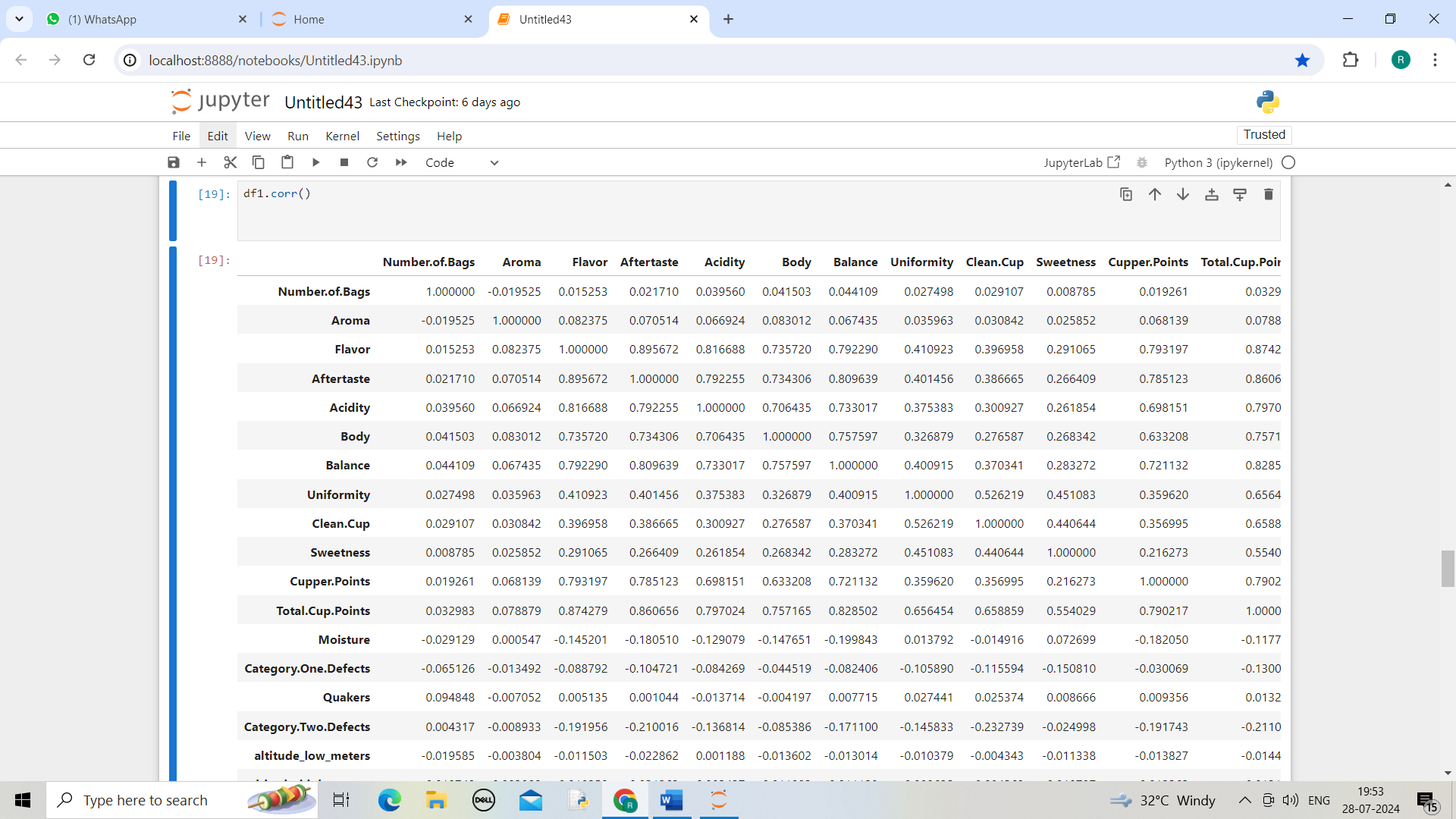
Strong Negative Correlation: Points cluster around a downward slope.

Weak or No Correlation: Points are scattered without a clear pattern.

7. Draw Conclusions

Based on your findings, make informed decisions about coffee quality. For example, if you find a strong positive correlation between roast level and flavor profile, you might adjust roasting processes to achieve desired flavors.

By following these steps, you can effectively perform bivariate analysis in coffee quality checking without coding, leveraging manual calculations and visualization tools.



Conclusion

To determine the conclusion of coffee quality using Python, you'd typically follow these steps:

Data Collection: Gather data on coffee attributes (e.g., aroma, acidity, body, flavor) and quality scores.

Data Preparation: Clean and preprocess the data to ensure it's suitable for analysis.

Feature Analysis: Identify key features that influence coffee quality. This could involve exploratory data analysis (EDA) to understand relationships between features and quality.

Model Building: Use statistical or machine learning models to analyze the data. This could include regression analysis, classification algorithms, or other techniques depending on your specific goal.

Evaluation: Assess the model's performance using metrics like accuracy, precision, recall, or RMSE (Root Mean Squared Error), depending on whether you're doing classification or regression.

Conclusion: Summarize the findings based on the analysis. For instance, you might determine which features most significantly impact coffee quality or predict quality based on certain attributes